

Statistical Considerations in Meta-Analysis

Hyun Kang

Department of Anesthesiology and Pain Medicine, Chung-Ang University College of Medicine, Seoul, Korea

The increase in medical research has led to a large body of related studies. The huge volume of research brings about a problem of how to organize and summarize the findings of studies. Meta-analysis is a statistical technique for combining the results from two or more studies, which addresses a similar hypothesis in a similar way. Meta-analysis includes the complete coverage of all relevant studies, and describes the results of each study via a quantitative index of effect size. Meta-analysis presents the precise estimate of treatment effect via combining these estimates across studies. Further, meta-analysis looks for the presence, degree and cause of heterogeneity, and explores the robustness of the main findings using statistical techniques. The author dealt with the some statistical issues and considerations which should be considered in conducting and presenting meta-analysis with explanation (ie. Effect size, Fixed and Random effect model, Heterogeneity, Reporting bias, and Meta-analysis Packages). This article may remind readers to conduct and evaluate the meta-analysis systematically and comprehensively.

Key Words: Research Design; Heterogeneity; Meta-Analysis; Review; Statistics

Correspondence to: Hyun Kang
우156-755, 서울시 동작구 흑석로 102,
중앙대학교 의과대학 마취통증의학과
Department of Anesthesiology and Pain
Medicine, Chung-Ang University College
of Medicine, 224-1 Heukseok-ro,
Dongjak-gu, Seoul 156-755, Korea
Tel: +82-2-6299-2571
Fax: +82-2-6299-2585
E-mail: roman00@naver.com

Received 1 December 2014

Revised 26 December 2014

Accepted 9 January 2015

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

서 론

의학분야에서는 환자의 생명을 연장하기 위하여 그리고 삶의 질을 향상시키기 위하여 많은 연구들이 시행되었으며, 시행되고 있다. 이 중 많은 연구들이 비슷한 질문에 답하기 위한, 비슷한 주제를 다루고 있다. 하지만 이 연구들이 항상 비슷한 결과를 보이는 것은 아니며, 오히려 상반된 결과를 보이는 경우도 있다. 연구들 중에는 특정한 중재가 유의하게 좋은 효과를 보인 경우, 중재가 좋은 효과가 있어 보이지만 통계학적으로 유의성에 이르지 못하는 경우, 중재가 오히려 나쁜 결과를 보인 연구 등도 있을 수 있다. 이럴 경우 우리는 연구의 결과들을 받아들이고, 현실에 적용하는 데 있어 어려움을 느끼게 될 것이다. 또한 어떤 주제에 있어서는 개별적인 연구들이 중재군에서 대조군에 비하여 좋은 효과가 있어 보이지만 적은 표

본수로 인하여, 통계학적으로 유의한 효과를 보이지 못할 수도 있다. 이 결과들을 적절한 통계학적 방법을 이용하여 통합할 경우 유의한 결과를 보일 수 있다.

메타분석(meta-analysis)은 두 개 이상의 개별적인 연구들의 추정치를 종합하여 요약 추정치(pooled estimate)를 합성하는 통계적 방법을 말한다. 다시 말해, 연구들에서 제시된 결과들의 통합된 요약 추정치를 정량적으로(quantitatively) 산출하여 효과 및 효율성을 평가하기 위해 사용되는 통계적 기법이다. 그렇기 때문에 메타분석을 분석들의 분석이라고 부르기도 한다. 체계적 문헌고찰(Systematic review)과 메타분석은 흔히 같은 의미로 혼용되어 사용되고 있지만, 이들은 명확한 차이가 있다. 체계적 문헌고찰은 특정한 연구 질문에 답하기 위하여 객관적이고 재현성이 확보된 방법론을 사용하여 확보 가능한 모든 연구들을 수집한 후, 이 연구들에서 나온 결과에 대

한 고찰과 분석을 실시하여, 그것들을 제시하는 것이다. 따라서 많은 체계적 문헌고찰에서 메타분석을 시행하고 있지만, 반드시 메타분석을 시행하여야 하는 것은 아니며, 메타분석을 시행할 수도, 시행하지 않을 수도 있다. 예를 들어 연구들 간의 특성이 너무 이질적이면 자료의 통합을 시도하지 않을 수 있다. 반대로 메타분석 또한 많은 경우에서 체계적 문헌고찰을 시행한 후 메타분석을 시행하지만, 체계적 문헌 고찰을 시행하지 않고도 메타분석을 시행할 수 있다[1,2].

메타분석과 체계적 문헌고찰을 시행할 경우 생기는 장점은 (1) 실험환경에 차이가 있는 독립적인 연구들을 종합하여 일반화시킬 수 있으며, (2) 많은 개별적인 연구결과를 근거로 가설을 검증하므로, 일부 연구결과에만 치우치지 않는 종합된 결과를 제시할 수 있으며, (3) 표본수가 증가하므로 단일 연구로부터 개발한 효과크기(effect size) 보다는 신뢰할만한 효과 추정치를 얻을 수 있으며, (4) 상이한 연구결과가 있을 때, 그 원인 규명이 가능하며, (5) 상반된 연구들 사이에서 발생하는 논쟁을 조절할 수 있다는 장점이 있다[3,4].

하지만, (1) 마치 사과와 오렌지를 한 데 섞는 것과 같이 서로 비교할 수 없는 다른 성질의 연구결과들을 종합하려는 것은 비논리적이며, (2) 타당성이 낮은 연구와 높은 연구의 결과를 구별하지 않고 그대로 종합하여 왜곡된 결론을 유도할 수 있으며, (3) 동일한 연구에서 여러 개의 결과를 얻었을 때 이 중 하나만 사용할 경우에는 정보를 상실할 수 있으며, 여러 개 모두 쓸 경우 비독립적인 자료를 독립적인 것처럼 사용할 수 있으며 그리고 (4) 종합할 연구를 수집할 때 대개 출판된 연구만을 표집 대상으로 하기 때문에 연구물 표집의 대표성이 문제될 수 있다는 비판을 받고 있다[5-7].

메타분석을 위해서는 통합하고자 하는 주제와 관련된 선행 연구물들의 수가 충분하여야 한다. 연구자는 연구자가 시행하고자 하는 연구에서의 질문과 분석 목적에 관련된 선행연구들을 조사하고 수집하며, 선행 연구들에서 제시한 정보들 중, 통합 가능한 정보들을 수집한 후, 종합하여 통합된 연구 결과를 제시한다. 연구결과를 종합하는 데 있어, 동일한 주제를 대상으로 시행된 연구들이라 할지라도 서로 일치하지 않은 척도를 가진 경우에는, 그 결과들을 요약하기 위하여 동일 척도로 바꾸어 주는 표준화 과정이 필요하게 되는데, 이러한 표준화된 척도를 효과크기(effect size)라 한다.

효과크기(effect size)

효과크기는 메타분석에서 연구의 결과를 요약하기 위해 사용되는 정량적인(quantitative) 지수이다. 즉 효과크기는 각 연구에서 관심을 가지고 있는 관계의 크기 혹은 강도를 반영한다. 어떤 의미에서, 효과크기는 모든 연구의 결과들이 쉽게 해석, 비교, 통합될 수 있도록, 표준화된 척도로 요약하는 것이다. 효과크기에는 많은 종류의 효과크기가 있으며, 메타분석을 시행할 때, 연구의 결과를 쉽게 분석할 수 있게 하고, 연구들 간 비교를 할 수 있도록 효과크기

를 선택하여야 한다.

먼저, 효과크기의 추정치(estimate)와 효과크기의 모수(parameter, 실제 효과크기)를 구별하는 것이 중요하다. 효과크기의 추정치는 표본을 대상으로 한 연구에서 구해진 표본에서의 효과크기로, 연구마다 얻어진 표본이 다르므로, 연구에서 제시된 효과크기가 약간씩 다를 수 있지만, 효과크기의 모수는 모집단에서의 효과크기로, 모집단은 고정되어 있으므로, 변하지 않는 실제 효과크기이다. 조사한 연구에서 얻을 수 있는 것은 효과크기의 추정치이며, 우리는 효과크기의 추정치를 이용하여, 효과크기의 모수에 대한 추정을 하게 된다. 만일 연구의 표본이 매우 크거나 표본의 변동이 무시할 수 있을 정도로 작을 경우에는, 효과크기의 추정치가 효과크기의 모수라고 가정할 수 있다.

효과크기의 선택은 개별 연구에서 사용된, 연구 디자인, 결과 측정 방식, 통계 적용 방식 등에 따라 달라질 수 있다. 연구에서 사용된 효과크기는 대부분, D family (표준화된 평균 차 등), Odds ratio family (승산비 등), r family (상관계수 등) 중 하나로 분류할 수 있다.

1. D family: 평균 차(mean difference)와 표준화된 평균 차(standardized mean difference)

많은 연구에서 치료 혹은 처치의 효과에 대한 결과를 연속형 척도로 제시하거나, 척도의 수가 많은 순서형 척도로 제시하게 된다. 이런 경우에는 평균 차(mean difference) 혹은 표준화된 평균 차(standardized mean difference)를 효과의 크기로 사용하게 된다. 동일한 측정도구로 측정한 경우에는 평균차이(mean difference, MD)를 이용하며, 다양한 측정도구로 측정한 경우에는 연구결과를 동일한 단위로 표준화한, 표준화된 평균 차(standardized mean difference, SMD)를 이용한다.

표준화된 평균 차는 치료군에서의 표본평균과 대조군에서의 표본평균의 차를 통합 표준편차로 나눈 값이다. 표준화된 표본평균 차의 공식은 다음과 같다.

$$d = \frac{\bar{X}^T - \bar{X}^C}{S}$$

\bar{X}^T : 치료군에서의 표본평균, \bar{X}^C : 대조군에서의 표본평균, S: 통합 표본 표준편차

여기에서 통합 표본 표준편차의 공식은 다음과 같다.

$$S = \sqrt{\frac{(n^T - 1)S_T^2 + (n^C - 1)S_C^2}{n^T + n^C - 2}}$$

n^T : 치료군의 표본 수, n^C : 대조군의 표본 수, S_T : 치료군의 표준편차, S_C : 대조군의 표준편차

여기에서 표준화된 표본 평균 차에 해당되는 표준화된 평균 차의 모수 값은 다음과 같다.

$$\delta = \frac{\mu^T - \mu^C}{\sigma}$$

μ^T : 치료군에서의 모평균, μ^C : 대조군에서의 모평균, σ : 통합 모 표준편차

이러한 표준화된 평균 차는 표준편차의 단위가 평균의 단위와 같으므로 쉽게 해석할 수 있으며, 다른 측정도구를 이용한 연구들에서도 동일한 의미를 갖는 것으로 해석할 수 있다[8].

표본 추출할 때 표준화된 평균 차에 생기는 불확실성(uncertainty)은 다음과 같은 분산(variance)의 공식으로 나타낼 수 있다.

$$V = \frac{n^T + n^C}{n^T n^C} + \frac{\delta^2}{2(n^T + n^C)}$$

n^T : 치료군의 표본 수, n^C : 대조군의 표본 수, δ : 표준화된 평균 차의 모수 값, 여기에서 표준화된 평균 차의 모수 값(δ)은 표준화된 표본 평균 차(d)로 대신할 수 있다. 이 공식에서 치료군과 대조군의 표본 수를 알아야, 분산을 계산할 수 있다는 점을 알아야 한다.

표준화된 평균 차는 정규분포를 따르는 경우가 많으며, 표준화된 평균 차가 정규분포를 따르고 있다면, 분산의 제곱근인 표준편차를 이용하여 효과크기의 모수 값(즉 실제 효과 크기)에 대한 신뢰구간을 계산할 수 있다.

효과크기의 모수 값(δ)에 대한 95% 신뢰구간은 다음과 같이 계산할 수 있다.

$$d - 1.96 \times \sqrt{V} \leq \delta \leq d + 1.96 \times \sqrt{V}$$

d: 표본에서 얻어진 표준화된 평균 차, V: 표본 추출에서의 표준화된 평균 차의 분산

표준화된 평균 차에 대한 몇 개의 다른 공식들이 효과크기의 모수 값(실제 평균 차)을 계산하는 데 사용할 수 있다[9].

2. Odds ratio family: 승산비(odds ratio), 상대위험도(relative risk), 위험차(risk difference), 효과 발현 필요 증례수(number needed to treat, NNT)

이산형 변수는 대개 승산비, 상대위험도, 위험차, 효과 발현 필요 증례 수 중 한가지로 보고하고 있다(Table 1)[10].

1) 상대위험도(risk ratio or relative risk): 두 군 간의 사건이 발생할 확률 간의 비율

Table 1에서 중재군에서 사건이 발생할 확률(p^T)은 $\frac{a}{a+b}$ 이며 대조군에서 사건이 발생할 확률(p^C)은 $\frac{c}{c+d}$ 이므로, 대조군에서 사건이 발생할 확률에 대한 중재군에서 사건이 발생할 확률의 비율(relative risk, RR)은 다음과 같다.

relative risk, RR)은 다음과 같다.

$$\frac{p^T}{p^C} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{(c+d) \times a}{(a+b) \times c}$$

2) 승산비(odds ratio): 두 군 간의 사건의 오즈에 대한 비율

Odds의 정의는 $\frac{\text{사건이 일어날 확률}}{\text{사건이 일어나지 않을 확률}}$ 이다. 여기에서 사건이 일어날 확률과 일어나지 않을 확률의 합은 1이므로, 사건이 일어날 확률을 p 라고 생각하면 사건이 일어나지 않을 확률은 $1-p$ 가 되며, 사건에 대한 odds는 $\frac{p}{1-p}$ 가 된다.

Table 1에서 중재군에서 사건이 일어날 확률(p^T)은 $\frac{a}{a+b}$ 이므로 중재군에서 사건에 대한 odds^T는 다음과 같다.

$$\frac{p^T}{1-p^T} = \frac{\frac{a}{a+b}}{1-\frac{a}{a+b}} = \frac{\frac{a}{a+b}}{\frac{(a+b)-a}{a+b}} = \frac{\frac{a}{a+b}}{\frac{b}{a+b}} = \frac{a}{b}$$

대조군에서 사건이 일어날 확률(p^C)은 $\frac{c}{c+d}$ 이므로, 대조군에서 사건에 대한 odds^C는 같은 요령으로 계산하면 $\frac{c}{d}$ 가 된다. 따라서 두 군 간의 사건의 odds의 비율(odds ratio, OR)은 다음과 같다.

$$\frac{\text{odds}^T}{\text{odds}^C} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{a \times d}{b \times c}$$

3) 위험 차(risk difference): 두 군 간의 사건이 발생할 확률의 차이

중재군에서 사건이 발생할 확률(p^T)은 $\frac{a}{a+b}$ 이며 대조군에서 사건이 발생할 확률(p^C)은 $\frac{c}{c+d}$ 이므로, 두 군 간의 사건이 발생할 확률에 대한 차이는 다음과 같다.

$$p^T - p^C = \frac{a}{a+b} - \frac{c}{c+d}$$

4) 효과 발현 필요 증례수(number needed to treat, NNT): 위해를 피하거나 편익을 발생시키기 위해 대조중재에 비해 시험중재를 더 받아야 하는 증례의 수

효과 발현 필요 증례수는 다음과 같다.

효과 발현 필요 증례수는 다음과 같다.

$$\frac{1}{|\text{Risk Difference}|} = \frac{1}{p^T - p^C} = \frac{1}{\frac{a}{a+b} - \frac{c}{c+d}}$$

이산형 변수에서는 표준화된 비율 차(risk difference), 로그 승산비 그리고 로그상대위험도 등으로 유효크기(effect size)를 정의한다. 로그 승산비를 예를 들어 설명하면, 표본에서의 로그 승산비는 다음과 같다.

$$\log(\text{OR}) = \log\left(\frac{\text{odds}^T}{\text{odds}^C}\right) = \log\left(\frac{\frac{p^T}{1-p^T}}{\frac{p^C}{1-p^C}}\right)$$

odds^T: 표본 치료군에서 사건이 발생할 오즈, odds^C: 표본 대조군에서 사건이 발생할 오즈, p^T : 표본 치료군에서 사건이 발생할 확률, p^C : 대조군에서 사건이 발생할 확률

표본 로그 승산비에 해당하는 로그 승산비의 모수 ω 값은 다음과 같다.

Table 1. Results presenting as dichotomous data

	Number of events	Number of non-events	Total
Experimental Group	a	b	a+b
Control Group	c	d	c+d
Total	a+c	b+d	a+b+c+d

$$\omega = \log\left(\frac{ODDS^T}{ODDS^C}\right) = \log\left(\frac{\frac{\pi^T}{1-\pi^T}}{\frac{\pi^C}{1-\pi^C}}\right)$$

$ODDS^T$: 모 치료군에서 사건이 발생할 오즈, $ODDS^C$: 모 대조군에서 사건이 발생할 오즈, π^T : 모 치료군에서 사건이 발생할 확률, π^C : 모 대조군에서 사건이 발생할 확률

로그 승산비는 연구 간에 비슷한 의미를 가지고 있기 때문에 결합하여 통합추정치를 만들더라도 적절한 것으로 알려져 있어서 이 분형 변수의 분석에서 자주 사용되고 있다[10].

표본추출할 때 로그 승산비에 생기는 불확실성은 다음과 같은 분산의 공식으로 나타낼 수 있다.

$$\begin{aligned} V &= \frac{1}{n^T p^T} + \frac{1}{n^T(1-p^T)} + \frac{1}{n^C p^C} + \frac{1}{n^C(1-p^C)} \\ &= \frac{1}{(a+b) \times \frac{a}{(a+b)}} + \frac{1}{(a+b) \times \frac{b}{(a+b)}} + \frac{1}{(c+d) \times \frac{c}{(c+d)}} + \frac{1}{(c+d) \times \frac{d}{(c+d)}} \\ &= \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \end{aligned}$$

n^T : 치료군의 표본 수, n^C : 대조군의 표본 수, p^T : 표본 치료군에서 사건이 발생할 확률, p^C : 표본 대조군에서 사건이 발생할 확률

표준화 평균 차의 경우와 같이, 로그 승산비도 대략 정규분포를 따르는 경우가 많으며, 분산의 제곱근인 표준편차를 이용하여 로그 승산비의 모수 값, 즉 실제 효과 크기에 대한 신뢰구간을 계산할 수 있다.

효과크기의 모수 값 ω 의 95% 신뢰구간은 다음과 같이 계산할 수 있다.

$$\log(OR) - 1.96 \times \sqrt{V} \leq \omega \leq \log(OR) + 1.96 \times \sqrt{V}$$

이분형 변수에서는 위험비, 위험차와 같은 것들이 효과 크기를 계산하는 데 사용할 수 있으며,

로그 위험비와 로그 위험차의 분산 공식은 다음과 같다.

$$\begin{aligned} V_{\log(\text{risk ratio})} &= \frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d} \\ V_{\log(\text{risk difference})} &= \frac{ab}{(a+b)^3} + \frac{cd}{(c+d)^3} \end{aligned}$$

표준화된 평균 차에 대한 몇 개의 다른 버전의 공식을 효과크기를 계산하는 데 사용할 수 있다. Odds ratio family의 로그 승산비, 로그 상대 위험도, 로그 상대 위험차는 대 표본에서 정규분포를 한 다는 가정을 이용한 정규근사를 이용하게 된다.

3. The r family: 상관계수(correlation coefficient)

많은 연구들에서 두 가지의 연속형 변수 사이의 관계를 측정하고 있으며, 상관계수를 이용하여 효과크기를 제시하고 있다. 이러한 상관계수에서 통계 분석을 수행하기 위하여 Fisher의 z-transformation을 시행하게 된다[11]. Fisher의 z-transformation 공식은

다음과 같다.

$$z = \frac{1}{2} \log\left(\frac{1+r}{1-r}\right)$$

r: 표본상관계수, z: 표본상관계수를 z-transformation한 값
표본 상관계수의 z-transformation에 해당하는 모상관계수의 z-transformation은 다음과 같다.

$$\zeta = \frac{1}{2} \log\left(\frac{1+\rho}{1-\rho}\right)$$

r에 상응하는 모상관계수는 모집단의 상관계수이며, z에 상응하는 ζ 는 ρ 의 Fisher z-transformation을 시행한 값이다.

상관계수를 z-transformation한 값은 대체로 정규분포를 따르며, z-transformation한 상관계수의 표본추출의 불확실성 또한 분산공식을 이용하여 표시할 수 있다.

$$V = \frac{1}{n-3}$$

n: 연구에서의 표본 수

효과크기의 모수 값 ρ 의 95% 신뢰구간은 다음과 같이 계산할 수 있다.

$$r - 1.96 \times \sqrt{V} \leq \rho \leq r + 1.96 \times \sqrt{V}$$

앞서 설명한 각 효과 크기는 상호 간에 전환이 가능하며, 이에 해당하는 공식들이 제시되고 있으나, 이를 적용할 때는 전환 여부에 대한 충분한 검토가 이루어져야만 한다.

고정효과모형(fixed effect model)과 변량효과모형(random effect model)

메타분석은 먼저 개별 연구의 요약추정치(summary estimate)를 계산하며, 각각의 연구에 가중치를 준 후, 가중치가 부여된 평균(weighted average)을 이용하여 각 연구들을 종합한 통합 요약통계 통계량을 산출하는 순서로 진행이 된다. 여기에서 가중치는 대개 표본 수의 크기에 근거하여 부여하게 되는데, 소규모의 연구는 대규모의 연구에 비해 우연에 의한 영향을 더 많이 받을 수 있기 때문에 요약 추정치를 계산하는 데 있어 대규모 연구에 대하여 상대적으로 더 많은 가중치를 부여하는 것이다[12]. 요약추정치의 모수 값을 추정하는 데 있어, 앞서 설명한 효과크기의 계수들 중 어떤 것을 사용하더라도 메타분석의 통계 방법은 유사하다.

메타분석에서 유효크기들을 통합하는 방법에는 동일한 모집단을 가정하여, 각 모수들의 변화량을 연구 내의 변동량에 의해서만 설명하는 고정효과모형(fixed effect model)과 연구 간의 변동량을 추가로 고려하는 변량효과모형(random effect model)이 있다. 이들 두 모형의 차이는 얻어진 요약추정치가 어떤 분포에서 얻어진 자료인지에 대한 통계학적 관점의 차이라고도 할 수 있다[8].

대개 메타분석은 먼저 연구 내의 변동만을 고려하는 고정효과모

형으로 분석을 실시한다. 하지만 동일하지 않은 실험 환경, 대상, 연구 방법에 의하여, 연구들 간의 동질성을 만족하지 않는 경우가 발생한다. 동질성을 만족하지 못하는 연구 결과들을 대상으로 연구 간 변동을 고려하지 않고 통합하게 되면 비뮌림(bias)을 유발할 수 있다. 이러한 이유로 유효크기의 모집단에 대한 동질성 검사를 실시하여야 하며, 동질성을 만족하지 못하면 연구 간의 변동을 고려한 변량효과모형을 가정하여 추론한다.

1. 고정효과모형(fixed effect model)

고정효과모형은 통합하고자 하는 연구들의 연구집단, 중재법이 동질하다고 판단되는 경우, 연구들의 수가 매우 적은 경우에 사용할 수 있다. 고정효과모형은 각 연구들에 있어 모집단은 동일하며, 중재 효과의 참 값은 단 하나만 존재(common true treatment effect)한다는 전제하에, 관찰된 치료 효과 값들의 차이는 표본 추출의 오차(sampling error) 때문이라는 가정에서 출발하는 모형이다(Fig. 1).

$$\theta_i = \mu + \epsilon_i$$

(θ_i : 표본 추출된 효과크기, μ : 효과크기의 모수와 ϵ_i : 표본 추출의 오차)

이때, 평균 θ_i 는 단 하나의 값만 가지고 있으므로 고정된 상수이며, ϵ_i 는 표본추출에 의하여 발생한 오차이다.

연구들의 효과크기 차이에 관련된 변동은 단지 i번째 연구의 연구 내 변동(within-study variation)인 $Var(\theta_i)$ 라고 가정하며, 효과 크기의 추정량 $\hat{\theta}_i$ 는 정규분포를 따르거나, 최소한 근사적으로 정규분포를 따른다고 가정한다.

그러면, 효과크기의 추정량 $\hat{\theta}_i$ 는

$$\hat{\theta}_i \sim N(\theta, Var(\hat{\theta}_i)), i = 1, 2, 3, \dots, k \text{ 로 표현할 수 있다.}$$

앞서 기술한 효과 크기들 중, 표준화된 평균 차와 Fisher z-transformed 상관계수에서는 정규성에 대한 가정이 거의 완벽하게 맞아 떨어지지만, z-transformation되지 않은 상관계수나 로그 승산비에서는 정규성에 대한 가정이 표본수가 아주 많이 커지지 않은 이상

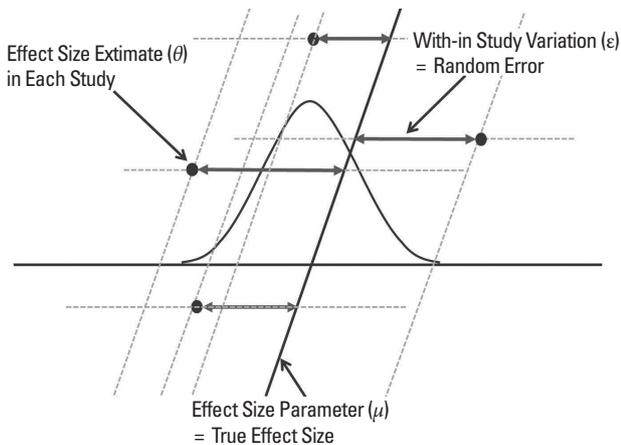


Fig. 1. Graphical presentation of fixed effect model.

성립되지 않는다.

개별 연구의 요약추정치(summary estimate)에 가중치를 부여한 평균(weighted average)을 이용하여 각 연구들을 종합한 통합 요약 통계량을 산출하게 되는데, 가중 평균은 다음의 식으로 계산이 가능하다.

$$\hat{\theta} = \frac{\sum_{i=1}^k \hat{\theta}_i w_i}{\sum_{i=1}^k w_i}$$

$\hat{\theta}$: 합성된 유효크기의 가중평균, $\hat{\theta}_i$: i번째 연구에서 추정된 효과 크기, w_i : i번째 연구의 가중치

고정효과모형에서 사용하는 방법은 역분산(generic inverse variance) 추정법, 멘텔-헨젤(Mantel-Haenszel) 추정법, Peto 추정법, 최대 우도(Maximum likelihood) 추정법 등이 있다[8].

1) 역분산 추정법 (generic inverse variance estimation method)

메타분석에서 가중치를 주는 데 있어 가장 많이 사용되는 방법으로, 효과 추정치 분산의 역수를 개별 연구의 가중치로 사용한다. 표본수가 큰 연구는 작은 분산을 가질 것이고, 분산의 역수는 커지게 될 것이므로, 표본수가 큰 연구에 더 큰 가중치를 주기 위해 사용하는 방법이다. 이 방법은 결합하고자 하는 연구들의 수는 적지만, 각 연구들의 표본수가 큰 연구들인 경우에 효과적인 방법이다.

역분산 추정법을 이용한 i번째 연구에서 추정된 효과크기 $\hat{\theta}_i$ 의 통합 추정량은 다음의 식으로 계산된다.

$$\hat{\theta} = \frac{\sum_{i=1}^k \hat{\theta}_i w_i}{\sum_{i=1}^k w_i} \sim N\left(\theta, \frac{1}{\sum_{i=1}^k w_i}\right)$$

여기에서 가중치인 $w_i = \frac{1}{SE^2}$ 이며, $N(\theta, \frac{1}{\sum_{i=1}^k w_i})$ 는 $\hat{\theta}$ 가 고정된 값을 가지며, 평균 θ 와 분산 $\frac{1}{\sum_{i=1}^k w_i}$ 인 정규분포를 한다고 가정한다는 것을 나타낸다.

2) 멘텔-헨젤 추정법(Mantel-Haenszel estimation method)

2 by 2 table을 만들 수 있는 자료에서 사용이 가능하며, 사건 발생률이 낮거나, 표본수가 작지만 연구의 수가 많을 때 효과적인 방법이다. 이 방법은 승산비나 상대위험도, 위험도 차에 대해 로그변환을 실시하지 않고 있는 값 자체를 사용한다.

$$\hat{\theta} = \frac{\sum_{i=1}^k \hat{\theta}_i w_i}{\sum_{i=1}^k w_i}$$

$\hat{\theta}$: 합성된 유효크기의 가중평균 $\hat{\theta}_i$: i번째 연구에서 추정된 효과 크기, w_i : i번째 연구의 가중치

θ 가 승산비일 경우 $\hat{\theta}_i$ 는 $\frac{axd}{bxc}$ 이며

여기에서 w_i 는 $\frac{1}{VAR_i}$ 이며 VAR1은 i번째 연구에서의 분산이다. 역분산법에서와 같이 승산비, 상대위험도, 위험 차에 대한 분산 공식은 다르다.

$$(1) \text{VAR(OR)} = \frac{(a+b) + (c+d)}{b \times c}$$

$$(2) \text{VAR(RR)} = \frac{(a+b) + (c+d)}{(a+b) \times c}$$

$$(3) \text{VAR(RR)} = \frac{(a+b) + (c+d)}{(a+b) \times (c+d)}$$

멘텔-헨젤 추정법에서는 공동 승산비(pooled odds ratio), 공동 상대위험도(pooled relative risk), 공동 위험차(pooled relative difference)를 추정하여 이용하게 된다.

3) Peto의 추정법

Peto의 방법은 승산비에서만 사용 가능하며, 이 경우 로그 승산비를 유효효과로 정의하고 추정하게 된다. 결과 수가 적거나, 사건 발생률이 낮을 경우(사건 발생률이 1% 미만), 그리고 실험군이나 대조군에서 결과가 발생하지 않은 경우(빈칸(zero cell)이 있는 경우)에 효과적인 방법이다. 이 방법은 치료효과 크기가 작을 경우(승산비가 1에 가까운), 개별 연구들이 각각 유사한 수의 실험군과 대조군을 가지고 있을 때 사용하게 된다. 집단 간 표본수에 불균형이 클 경우에는 효과를 과대 추정하게 하며, 실제 효과 크기가 큰 경우에는 과소추정을 하는 경향이 있다. 따라서, 두 군 간의 대상 환자 수가 매우 다르거나, 효과 크기가 매우 크거나, 사건 발생률이 5% 이상이라면 멘텔-헨젤 방법을 사용하는 편이 나올 수 있다.

2. 변량효과모형(random-effects model, derSimonian-laird estimation method)

변량효과모형은 각 연구들에 있어 중재효과의 참값은 단 하나만 존재하는 것이 아니라 각 연구들에 있어 중재법의 효과는 어떤 중재효과 평균 참값(true average treatment effect)을 중심으로 정규분포를 따른다고 가정하는 모형이다. 즉, 각 연구들은 중재효과 평균 참값을 중심으로 퍼져있는 모 집단 내 연구들로부터 무작위로 추출된 연구들이고, 이 추출된 연구들의 평균값은 평균 참값을 중심으로 정규분포를 따른다는 의미이다(Fig. 2).

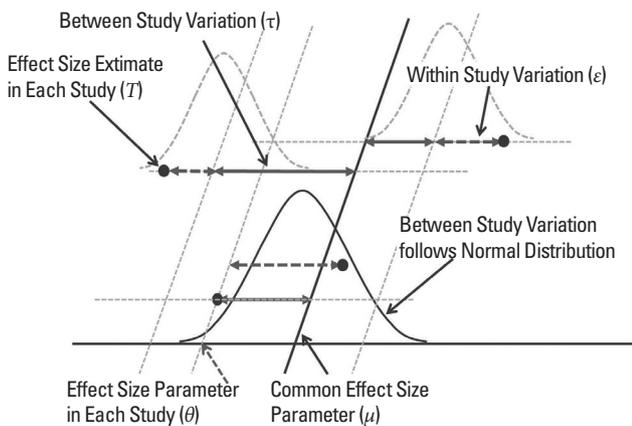


Fig. 2. Graphical presentation of random effect model.

변량효과모형은 다음과 같이 표시할 수 있다.

$$T_i = \theta_i + \epsilon_i = \mu + \tau_i + \epsilon_i$$

(T_i : 진실된 중재효과 θ_i 과 변동 ϵ_i 을 가진 분포로부터 표본 추출된 효과크기. 여기서 중재효과의 참값 θ_i 은 평균 μ 와 변동 τ_i 을 가진 분포로부터 표본 추출된 것임)

따라서 각 연구의 중재효과들 간에 관찰되는 변동(variation)은 각 연구들 간의 변동(between-study variation, τ)과 표본추출의 오차(within-study variation: random variation, ϵ)를 포함한 것이라고 가정한다. 그러므로, 자료 통합 시 연구들 간의 변동과 표본추출의 오차를 모두 고려한 가중치를 사용하여야 한다.

변량효과모형에서의 가중치

$$\text{weight} = \frac{1}{\text{variance}_{\text{within study}} + \text{variance}_{\text{between study}}(\tau^2)}$$

연구 간 변동의 추정치가 0인 경우(연구 간 이질성이 없는 경우) 변량효과모형은 고정효과모형과 동일한 추정치를 제공한다. 하지만 연구 간의 이질성이 심할 때 변량효과모형은 더 많은 변동을 인정하기 때문에, 고정효과모형에 비해 치료 효과의 유의성에 대해 좀더 보수적인 추정치(더 넓은 신뢰구간)를 제공한다. 변량효과모형은 고정효과모형에 비해 작은 규모의 연구에 더 높은 가중치를 주고, 더 큰 규모의 연구에는 더 낮은 가중치를 주기 때문에 치료 효과의 추정치는 두 모형 간에 다를 수 있다.

변량효과모형에서 각 연구들은 동일한 중재효과를 추정하기 보다는 다른 중재효과를 추정하며, 다음과 같은 경우에 변량효과모형을 고려할 수 있다.

- (1) 메타분석에 포함된 연구의 인구집단, 중재법 등이 동질하지 않다고 판단될 때
 - (2) 메타분석의 목적이 일반적으로 폭넓은 인구집단에서 일반화시키는 것일 때
- (예, 연구대상이나 중재법의 다양성으로 하나의 치료효과 크기를 가정하기 어려운 경우)

변량효과모형에서의 가중치는 연구 간 변이(between studies variation, τ)가 포함된다라는 것이 고정효과모형과 다른 점이며 어떤 추정방법을 사용하느냐에 따라 연구 간 변이를 구하는 방법이 달라진다. 이를 위한 통계적 방법으로는 가중최소제곱(weighted least squares) 추정법, 비가중최소제곱(unweighted least squares) 추정법, 최대우도(maximum likelihood) 추정법 등이 있다.

변량효과모형에서 가중치를 추정할 때 일반적으로 사용되는 방법은 가중최소제곱추정법인 Dersimonian and Laird추정법이다 [13].

$$\text{weight} = \frac{1}{D + \frac{1}{w_i}} \quad w_i = \frac{1}{\text{variance}_i}$$

이질성이 있을 경우 변량효과모형을 사용하게 되지만, 변량효과모형이 이질성 문제를 해결하기 위한 이상적인 방법은 아니다. 변량

효과모형은 비뿔림이나 교란변수를 보정해 주지 않으며, 이질성(heterogeneity)을 일으키는 원인을 해결해주는 것도 아니기 때문에 연구 간 이질성이 심할 경우 변량효과모형으로부터 계산된 효과 추정치는 부적절할 수 있다. 그러므로 변량효과모형이 이질성을 설명하기 위한 것으로 대체되어 사용되어서는 안 된다. 연구들 간에 나타난 이질성 정도가 지나치게 클 경우에는 변량효과모형을 이용한 통합 방법을 사용하지 않고, 결과를 따로 제시하는 것이 더 좋은 방법이 될 수 있다.

이질성(heterogeneity)

메타분석을 시행하는 목적은 요약추정치(summary estimate)를 알아내는 것만이 아니며, 연구 결과에서의 양상을 알아내는 것 또한 중요하다. 연구 결과의 양상이 일관되게 나타난다면 그 의미가 무엇인지, 일관되지 않게 나타난다면, 그 의미가 무엇인지, 일관되지 않게 만드는 원인을 파악하는 것이 중요하다. 여기서 일관되지 않게 나타나는 성질을 이질성이라고 하며, 메타분석에서의 이질성이란 메타분석에서 통합되는 개별연구들의 결과에서 나타나는 변이(variation)가 표본추출에서 생기는 변이 이상으로 관찰되어, 변이가 우연에 의하여 발생하였다고 설명할 수 없는 경우를 의미한다. 이질성은 임상적 다양성, 방법론적 다양성, 우연(chance), 비뿔림에 의해 발생할 수 있다. 임상적 다양성은 연구에서 대상이 된 환자나 집단들의 연령, 성별, 질병의 중증도, 진단명 등과 같은 성격, 연구에서 사용된 중재법, 병용중재법, 결과를 측정하기 위해 사용된 방법 및 척도의 차이, 추적조사 기간 등에 의해 발생할 수 있으며, 방법론적 다양성은 사용된 연구의 유형과 배정순서은폐, 눈가림과 같은 비뿔림을 일으킬 수 있는 요소를 포함하는 연구방법에서의 질적 차이 등에 의하여 발생할 수 있다.

이질성은 또한 우연에 의해 발생할 수 있는데, 표본크기가 작은 연구나 사건 발생률이 낮은 연구들은 표본 추출 시 우연히 효과크게 나타날 수 있다. 이질성은 연구비 출처나 출판 등에 의한 비뿔림에 의해서도 발생할 수도 있다.

이질성의 존재 여부 및 정도는 그래프를 이용한 시각적 검사나, 통계적 검정에 의하여 확인할 수 있다. 그래프를 이용한 시각적 검사방법에서 가장 대표적인 방법은 숲그림(Forest plot)을 그린 후 개별 연구들의 치료효과 값의 방향성과 신뢰구간이 겹치는지 확인함으로써 이질성을 확인하는 방법이다. 만일 신뢰구간이 겹친다면, 이질성이 적은 것으로 판정하며, 신뢰구간이 겹치는 부분이 적게 된다면, 추가적으로 통계적 검정방법을 통하여 이질성 여부 및 정도를 확인하여야 한다.

그래프를 통한 이질성 여부 및 정도를 확인하는 또 다른 방법으로 L'Abbe Plot과 Galbraith plot을 이용하는 방법이 있다. L'Abbe Plot은 이항변수 자료에 대하여 대조 집단과 중재 집단의 확률을 통합과정에서 사용되는 가중치가 고려된 원들을 X-Y 평면에 표시

하여 시각적으로 확인하는 방법으로 연구들이 원점에 가까운 구름 모양의 직선 형태이면 동질성을 만족한다고 판정한다. Galbraith plot 효과크기의 표준오차의 역수와 표준화된 효과크기를 X-Y 평면에 표시하여 시각적으로 확인하는 방법이다. 원점을 지나 는 회귀선을 추정하고 회귀선을 기준으로 ±2 안에 점들이 들어 있으면, 동질성을 만족한다고 판정한다. 하지만 이 두 방법 모두 흔히 사용되지는 않는다.

이질성을 확인하기 위한 통계적 검정방법으로는 카이제곱 검정법(Q statistics), Higgin's I² statistic 등이 있다.

1. 카이제곱 검정법(Q statistic)

이질성 검정을 위한 카이제곱 검정은 Q 통계량이라고도 하며, 각 연구들의 중재효과가 공통중재 효과추정치로부터 얼마나 멀리 떨어져 있는지 검정하는 방법이다.

Cochrane의 Q 통계량은 k개의 연구 결과들이 동질하다는 귀무가설 $H_0 : W_1 = W_2 = \dots = W_k$ 을 검정하게 된다.

Q 통계량은 다음과 같은 공식으로부터 계산이 가능하다.

$$Q = \sum_{i=1}^k W_i (Y_i - M)^2 \sim \chi^2 (k - 1)$$

여기에서 $W_i = \frac{1}{V_i} = \frac{1}{S_i^2}$ 이므로

$$\begin{aligned} &= \sum_{i=1}^k \left(\frac{Y_i - M}{S_i} \right)^2 \\ &= \sum_{i=1}^k \left(\left(\frac{Y_i}{S_i} \right)^2 - 2 \frac{Y_i M}{S_i} + \left(\frac{M}{S_i} \right)^2 \right) \\ &= \sum_{i=1}^k \frac{Y_i^2}{S_i^2} - 2 \sum_{i=1}^k \frac{Y_i M}{S_i} + \sum_{i=1}^k \frac{M^2}{S_i^2} \\ &= \sum_{i=1}^k W_i Y_i^2 - \frac{\left(\sum_{i=1}^k W_i Y_i \right)^2}{\sum_{i=1}^k W_i} \end{aligned}$$

W_i : i번째 연구의 가중치, Y_i : i번째 연구의 효과크기, V_i : i번째 연구의 분산, S_i : i번째 연구의 표준편차, M : 요약추정치(공통 중재효과)

만일 Q값이 크면 M의 변동성이 크다는 것을 의미하므로 귀무가설 $H_0 : W_1 = W_2 = \dots = W_k$ 를 기각하여 연구결과가 서로 동질적이지 않다고 결론을 내린다.

카이제곱 검정은 χ^2 분포를 이용하여, P값을 제시해 주므로, 객관적인 판단을 할 수 있게 해주기는 하지만, 효과크기에 민감하지 않고, 연구의 수에 민감한 특징을 가진다. 따라서 연구들이 작은 규모이거나 포함된 연구의 수가 적을 때 낮은 검정력을 가지며, 연구수가 많을 때는 높은 검정력을 가지게 된다. 메타분석에서는 연구의 수가 적으므로, 낮은 검정력을 가지게 되어, 중요하지 않은 이질성을 발견해 낼 가능성이 증가하게 되므로, 해석에 주의를 기울여야 한다. 그러므로 낮은 검정력을 가질 수 있다는 점을 고려하여, 0.05 대신 0.10을 통계적 유의수준으로 사용한다[14].

2. Higgin's I^2 statistic

Higgin's I^2 통계량은 이질성의 정도를 정량화시킨 통계량이다 [15].

$$I^2 = \frac{Q - df}{Q} \times 100\%$$

Q: 카이제곱 통계량(Q-statistic), df: Q statistic의 자유도

I^2 는 일종의 noise ratio로 관찰 효과의 전체 변동량 중 이질성으로 인한 변동량의 비율이다. I^2 는 Q 통계량과는 달리, 척도나 연구의 수 모두에 민감하지 않은 특징을 가진다.

I^2 는 0%에서 100% 사이의 값을 가지게 되며, 이질성이 없다면 0%가 되며, 이질성이 증가함에 따라 I^2 값이 증가하게 된다.

여러 가지 해석 기준이 존재하고 있으나, 일반적으로 해석을 위해 제시된 기준은 다음과 같다.

0% ≤ I^2 ≤ 25%: 이질성이 낮음

25% ≤ I^2 ≤ 75%: 중간 정도의 이질성이 있을 수 있음

75% ≤ I^2 ≤ 100%: 상당한 이질성이 있음

보통 고정효과모형과 변량효과모형의 선택은 $I^2 = 50\%$ 를 기준으로 하게 된다.

이질성이 있는 경우의 메타분석

메타 분석을 실시하는 도중에 이질성이 확인되었다면, 먼저 문헌 선정에서부터 입력된 자료들이 정확한지 다시 한번 확인하여야 한다. 메타분석의 기본적인 가정인 동질성을 충족시키지 못한다고 판단이 된다면 메타분석을 시행하지 않고, 체계적 문헌 고찰만을 실시할 수 있다.

어느 정도 이질성이 있는 경우, 이질성이 있는 연구들의 통합을 위하여 변량효과 모형을 사용할 수 있다. 이질성이 있는 연구들에서 변량효과 모형은 고정효과 모형에 비해 상대적으로 작은 규모의 연구들에 더 많은 가중치를 준다. 변량효과 모형의 분산은 고정효과 모형에서의 표본 추출 오차에 추가적으로 연구 간 변동을 포함한다. 따라서 변량효과 모형에서의 변동이 커지게 되며, 요약추정치의 신뢰구간은 고정효과 모형에서 보다 상대적으로 넓어지게 된다.

또한 이질성이 존재할 경우 이질성의 원인을 찾아볼 수 있는데, 이를 위해서는 하위그룹분석이나, 메타회귀분석, 민감도분석을 실시할 수 있다[16].

1. 하위그룹분석(subgroup analysis)

연구들 간의 이질성을 해석하고 이질성의 원인을 분석하기 위해서는, 동질 할 것으로 기대되는 하위그룹을 나눈 후 하위그룹별로 분석한 후 이를 비교하여 원인을 분석해 볼 수 있다. 하위그룹분석의 목적은 연구들 간의 이질성이 임상적인 특성이나 방법론적인 특성에 기인하는지 파악하고자 하는 것이다. 임상적인 특성에는 포

함, 배제 기준, 기저질환, 인구학적 특성, 진단기준, 결과 변수에 대한 정의 등이 포함되며, 방법론적인 이질성에는 연구설계의 차이, 연구의 질의 차이 등에 기인하는지 파악하는 것이다.

무작위대조군 연구에서 하위그룹분석을 연구 전에 사전에 계획 하라고 제시되는 것처럼, 메타분석에서도 하위그룹분석은 사전에 계획되어야 한다. 하지만 분석을 수행할 변수를 사전에 모두 알기는 어렵기 때문에 이질성을 확인한 후 몇몇 변수에 대해서 하위그룹분석을 할 수는 있으나, 사후 하위그룹분석을 시행할 경우 해석에 많은 주의를 기울여야 한다. 또한 하부그룹분석을 시행할 경우 연구의 수가 더욱 작아지기 때문에 검정력이 낮아지게 되며, 이로 인해 해석에 더욱 주의를 기울여야만 한다. 사전에 구체화된 특성들에서 하위그룹의 수를 제한하여 통계적인 오류의 가능성을 줄여야 하며, 하위그룹이 분석되는 데 영향을 미치는 연구 결과들을 연구자가 모르게 하여 비뚤림이 있는 결과를 이끌어낼 가능성을 줄여야 한다.

2. 메타회귀분석(meta-regression)

메타회귀분석은 개별 연구들에서 하나 혹은 그 이상의 공변량에 대한 연구 요약추정치의 회귀분석이다. 공변량이 개개의 대상의 수준이 아니라 연구 수준이라는 점을 제외하면 개별 연구에서 이용하는 회귀분석을 메타분석에 적용하였다고 생각해 볼 수 있다. 치료효과와 조사된 연구 특성 간에 선형관계가 있다면 메타회귀분석은 하위그룹분석보다 통계적 검정력이 더 크다[17].

회귀분석에서 종속변수인 y축에는 중재효과크기인 로그 변환된 승산비(lnOR), 표준화된 평균차(SMD), 가중평균차(WMD), 로그 변환된 상대위험도(lnRR), 로그 변환된 hazard ratio (lnHR) 등이 되며, 설명변수인 x축은 중재효과크기에 영향을 미칠 수 있는 공변량인 연구들의 특성이 된다.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \epsilon$$

y_i : i번째 연구의 효과크기, β_0 : 추정된 전체 효과크기, x_i : 중재 효과크기에 영향을 미칠 수 있는 연구의 특성, ϵ : 연구 간 변동

연구들 간의 모든 변동은 회귀모형에 포함되는 공변량으로 설명이 가능하다고 생각될 때 고정효과 메타회귀분석을 사용하는 것이 적절하며, 변량효과 메타회귀분석은 공변량에 의해 설명되지 않는 연구들 간의 변동인 잔여 이질성(residual heterogeneity)을 허용하므로, 공변량으로 설명이 가능하지 않을 경우 변량효과 메타회귀분석을 사용하는 것이 적절하다. 그러나 이질성의 원인이 모두 설명이 가능하다고 가정하는 것은 합리적이지 않으므로, 연구 수준의 인자에 의해 설명되지 않는 잔여 이질성(residual heterogeneity)을 허용하기 위해서는 고정효과 메타회귀분석 대신 변량효과 메타회귀분석을 수행하는 것이 합리적이다. 만일 단변량 메타회귀분석을 여러 번 나누어 실시하게 되거나 다변량 메타회귀분석을 여러 개의 공변량을 이용하여 실시하게 되면, 위양성 확률을 증가시

켜 1종 오류의 가능성을 높일 수 있다.

3. 민감도분석(sensitivity analysis)

민감도 분석은 메타분석의 결과가 각 단계에서 행해진 결정사항이나, 가정들을 변화시킴으로써 쉽게 변화되는지를 파악하고자 하는 분석이다[2]. 가정들에서의 결정사항이나, 가정들을 변화시켰을 때에도 결과의 변화가 민감하지 않다면(robust) 우리는 그 결과를 더욱 신뢰할 수 있기 때문이다. 출판되지 않은 연구, 질이 낮은 연구, 결측 값 등을 제외시킨 후 결과들이 변하는지 확인함으로써, 민감도 분석을 행하게 된다.

보고비뚤림(reporting bias)

보고비뚤림은 연구 결과의 보고 및 확산이 결과의 특성과 방향에 의해 영향을 받을 때 발생하는 비뚤림으로서, 이에는 출판 비뚤림(publication bias), 언어 비뚤림(language bias), 위치 비뚤림(Location bias), 시간지연 비뚤림(time lag bias) 등이 있다. 이중 가장 대표적인 것은 출판 비뚤림(publication bias)으로, 이는 통계적으로 유의한 차이를 보이는 연구가 통계적으로 차이가 없는 연구보다 출판될 가능성이 높게 되며, 따라서 메타분석을 위한 문헌 검색 시에도 통계적으로 유의한 결과를 보인 연구가 더 많이 파악되어, 메타분석결과가 왜곡되어 발생하는 비뚤림이다[18]. 비뚤림은 연구결과를 과대추정 혹은 과소 추정하게 만들며, 잘못된 연구결과를 야기하게 되어, 잘못된 해석이나 해석에 있어 어려움을 일으키게 된다.

보고비뚤림의 존재나 정도를 추정하는 방법에는 그래프를 이용한 방법과 통계적 검정방법이 있다.

그래프를 이용하는 방법 중 가장 대표적인 방법은 깔때기 그림(Funnel plot)을 이용하는 것이다. 깔때기 그림은 Y축에 연구의 표본크기를, X축에 효과크기를 위치시킨 후, 각 연구들을 점으로 표현하는 산포도로, 위쪽이 좁으며 아래쪽은 넓은 깔때기를 엮어놓은 모양을 가지게 된다.

보통 Y축은 $\frac{1}{\text{표준오차}}$ 로 표시하며, 연구에서 표본수가 증가하게 되면, 표준오차는 감소하게 될 것이며, $\frac{1}{\text{표준오차}}$ 은 증가하게 될 것이다. 그러므로 Y축의 상단은 표본수가 큰 연구, 다시 말해 대규모의 연구가 위치하게 된다. X축은 효과크기가 되며, 오른쪽으로 갈수록 효과크기가 증가하게 되고, 왼쪽으로 갈수록 효과크기가 감소하게 된다. 깔때기 그림을 시각적으로 판단하여 대칭 여부에 따라 보고비뚤림 여부를 판단하게 되는데, 만일 엮어진 깔때기의 좌측 하단이 비어있게 된다면, 소규모 연구로 효과크기가 적은 연구가 출판되지 않았음을 의미하며, 우측 하단이 비어있게 된다면, 소규모 연구로 효과크기가 큰 연구가 출판되지 않았음을 의미한다. 이는 보고비뚤림을 시사하는 소견이다. Contour-enhanced funnel plot은 깔때기 그림에 통계적 유의성을 나타내는 선을 추가하여 제시하여

주는 방법이다[19].

통계적 검정을 이용하는 방법으로는 Begg and Mazumbar's rank correlation test [20], Egger's test [21] 등이 있다.

Egger's test의 검정은 중재 효과 값의 역수에 가중치가 부여된, 표준오차에 대한 중재효과 추정치의 선형 회귀검정으로 Y 절편이 0과 유의하게 다르지 평가한다.

만일 보고비뚤림이 있을 경우, 비뚤림이 중재효과에 미치는 영향을 확인하여, 보정을 하여 중재효과에 변화가 있는지 확인할 수 있다. 이중 가장 대표적인 방법은 절삭과 채움 방법(Trim and fill method)이다[22]. 절삭과 채움 방법의 순서는 먼저 절삭되거나 보고되지 않았을 것으로 생각되는 연구들의 수를 추정한 다음, 이 연구들을 절삭(trimming)한다. 다음으로, 대칭인 연구들을 대상으로 메타분석을 실시하여 중재효과에 대한 추정치를 계산한다. 비대칭성이 남아있지 않을 때까지 이 과정을 반복한다. 그리하여 요약통계량을 계산한다.

메타분석 관련 프로그램

메타분석을 수행하기 위한 프로그램들은 많이 있다. 그 중에서 무료로 제공되면서 메타분석 전문프로그램인 RevMan (<http://www.cc-ims.net/RevMan>), MIX (<http://www.meta-analysis-made-easy.com/download/>)과 R (<http://cran.r-project.org/bin/windows/base/>) 등이 있으며, 유료이면서 메타분석 전문프로그램인 Comprehensive Meta-Analysis Software (Version5.0, <http://www.meta-analysis.com>), 그리고 상용 통계분석전문 소프트웨어 이면서 고급화된 메타분석 수행이 가능한 STATA (<http://www.stata.com>), SAS 등을 들 수 있다.

결 론

메타분석은 비슷한 질문에 연구들에서 근거를 합성하는 체계적이며, 양적인 방법을 제공한다. 메타분석은 독립적인 연구들을 종합하여 일반화시킬 수 있으며, 일부 연구결과에만 치우치지 않는 종합된 결과를 제시할 수 있으며, 또한 신뢰할만한 효과 추정치를 얻을 수 있으며, 상이한 연구결과가 있을 때, 그 원인 규명이 가능하다는 장점이 있다. 현재 메타분석으로 시행된 연구들이 증가하고 있으며, 이 영역에서 급속히 발전이 이루어지고 있다. 그렇기 때문에, 연구자들은 메타분석의 장점과 단점, 그리고 메타분석에서 시행되고 있는 방법론, 특히 통계적 방법에서의 이슈에 대한 이해도를 높이고, 친숙해져야만 할 것이다. 메타분석에서 제시하는 결과를 무비판적으로 받아들이거나, 임상에 적용해서는 안될 것이며, 메타분석에 대한 지식 및 경험, 임상적 근거 등을 통합하며, 비판적으로 수용할 수 있는 역량을 키워나가야 하겠다.

REFERENCES

1. Greenhalgh T. How to read a paper: papers that summarise other papers (systematic reviews and meta-analyses). *BMJ* 1997;315:672-5.
2. Higgins JP, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from: <http://www.cochrane-handbook.org>.
3. Wolf FM. *Meta-Analysis: Quantitative Method for Research Synthesis*. Beverly Hills, CA: SAGE Publications, Inc; 1986.
4. Pillemer DB, Light RJ. Synthesizing outcomes: how to use research evidence from many studies. *Harvard Educ Rev* 1980;50:176-95.
5. Egger M, Smith GD. Bias in location and selection of studies. *BMJ* 1998;316:61-6.
6. Egger M, Smith GD, Altman DG. *Systematic reviews in health care: meta-analysis in context*. London: BMJ Publishing Group; 2001.
7. Furberg CD, Morgan TM. Lessons from overviews of cardiovascular trials. *Stat Med* 1987;6:295-306.
8. Hedges LV, Olkin I. *Statistical methods for meta-analysis*. Academic Press; 1985.
9. Rosenthal, R. Parametric measures of effect size. In Cooper H, Hedges LV, eds. *The handbook of research synthesis*. New York: Russell Sage Foundation; 1994;231-44.
10. Baumeister RF. Self-regulation, ego depletion, and inhibition. *Neuropsychologia* 2014;65:313-9.
11. Fisher RA. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* 1915; 10:507-21.
12. Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. *BMJ* 1997;315:1533.
13. DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: an update. *Contemp Clin Trials* 2007;28:105-14.
14. Fleiss JL. Analysis of data from multiclinic trials. *Control Clin Trials* 1986;7:267-75.
15. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557-60.
16. Song F, Sheldon TA, Sutton AJ, Abrams KR, Jones DR. Methods for exploring heterogeneity in meta-analysis. *Eval Health Prof* 2001;24:126-51.
17. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002;21:1559-73.
18. Simes RJ. Confronting publication bias: a cohort design for meta-analysis. *Stat Med* 1987;6:11-29.
19. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *J Clin Epidemiol* 2008;61:991-6.
20. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994;50:1088-101.
21. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629-34.
22. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000;56:455-63.